Online Supplementary Materials for Tsunemoto, A. & Trofimovich, P. (in press). Coherence and comprehensibility in second language speakers’ academic speaking performance. *Studies in Second Language Acquisition.*

## 

## Appendix A

Development of Study-Specific Rubric for Coherence Rating

Three pilot participants contributed to the development of the study-specific definition of coherence. These participants, who came from the same population as the EAP instructors in the main study, were three female instructors (*Mage* = 46.33 years, *SD* = 10.97), including two L1 English speakers and one English–Turkish bilingual. They had lived in Canada for about 22 years and had graduate degrees in applied linguistics or TESOL and an average of 16 years of experience teaching EAP speaking skills in higher education.

Because raters’ understanding of coherence can depend on various contextual details, including raters’ background and experience, the demands of a speaking task, and the context of assessment (Richards, 1990), it was important to develop a custom-based, study-specific definition of coherence. Therefore, three representative EAP instructors participated in a pilot project involving an individual interview via Zoom, where they rated and commented on samples drawn from the 60 target audio recordings. Although there is no publicly available or widely accepted definition of coherence, as it applies to integrated speaking task performances, following Iwashita and Vasquez (2015), coherence was initially defined for the instructors as how much spoken discourse is interrelated, unified, and meaningful to understand.

Raters first read and signed the consent form, then after familiarizing themselves with the task and the two source texts and using an additional audio for practice, the instructors evaluated four audio samples, each representing one of the target L1 groups (Farsi, Indian, Mandarin, Romance). The audios were purposefully selected to illustrate performances varying in the quality of discourse structure (e.g., with and without an introduction or a summary) and features (e.g., with and without signposting), including the targeted cohesive devices. After listening to each audio, the instructors first summarized their initial impressions about coherence. They then listened to the same audio again and explained specific reasons for their evaluation by stopping the recording to indicate what they were thinking about at each chosen location, pointing out reasons that impacted their decision to upgrade or downgrade the speaker’s coherence. Based on the instructors’ comments, the final definition of coherence for the main study was developed (with specific examples available in the next page):

Coherence refers to how well a speaker makes links and expresses relationships between different ideas. If ideas are clearly related to each other, if they are logically connected, and there is no missing information, then a speaker is highly coherent. However, if ideas are not well connected, if they are presented out of sequence, and if you have to fill in gaps to piece together missing information, then a speaker is not coherent.

Definitions and examples of coherent and comprehensible performance

**Comprehensibility**

This term refers to how much effort it takes to understand what someone is saying. If you can understand with ease, then a speaker is highly comprehensible. However, if you struggle and must listen very carefully, or in fact cannot understand what is being said at all, then a speaker has low comprehensibility.

|  |
| --- |
| This measure will be rated using this scale:  ⎔-------------------------------------------------------------------⎔  hard to understand easy to understand |

Here are some examples of how speakers can make their speech easy to understand:

* A speaker can use intonation to make their speech sound lively and engaging.
* A speaker can pace their speech so that it’s not too fast or too slow.
* A speaker can use varied, appropriate words to convey ideas.
* A speaker can use different grammar structures to convey precise, clear meanings.
* Even if there are problems with pronunciation, pausing or hesitations, vocabulary, or grammar, these problems don’t make the speaker’s message harder to understand.

**Coherence**

This term refers to how well a speaker makes links and expresses relationships between different ideas. If ideas are clearly related to each other, if they are logically connected, and there is no missing information, then a speaker is highly coherent. However, if ideas are not well connected, if they are presented out of sequence, and if you have to fill in gaps to piece together missing information, then a speaker is not coherent.

|  |
| --- |
| This measure will be rated using this scale:  ⎔-------------------------------------------------------------------⎔  not coherent at all very coherent |

Here are some examples of how speakers can make their speech coherent:

* A speaker can state a general idea and then illustrate it clearly with specific examples.
* A speaker can use specific expressions (e.g., *first*, *finally*, *to summarize*, *in contrast*) to show how ideas are structured.
* A speaker can illustrate how ideas are connected by explaining links between them.
* A speaker can use connectives (e.g., *and*, *but*, *because*, *despite*, *although*) to indicate how different utterances are linked.
* If a speaker uses pronouns (e.g., *he*, *she*, *they*, *it*, *one*), it is clear which person(s) or object(s) are being discussed.

**Appendix B**

Speech Prompt

**Integrated Task Academic: Version A**

Topic: psychology (actor-observer)

Reading source: participants read a 105-word text describing the actor-observer bias, wherein people form attributions differently depending on who performed an action. While individuals attribute the behavior of others based on internal factors such as their personality, they attribute their own behavior to external factors such as situations.

Listening source (lecture): after reading the text, participants listened to a lecture (3 min 15 sec) in which a professor provided two examples illustrating the actor-observer bias.

Question: participants were asked to respond to a question regarding how the two examples provided in the lecture could illustrate the actor-observer bias.

Reference:

Reading source:

Educational Testing Service. (2006). *The official guide to the new TOEFL iBT*. McGraw Hill. p. 288

Listening source:

Educational Testing Service. (2006). *The official guide to the new TOEFL iBT*. McGraw Hill. Disc 2: Track 58

**Integrated Task Academic: Version B**

Topic: sociology (social interaction)

Reading source: participants read a 93-word text describing the audience effects, which suggest that people behave differently when they are aware that they are being observed by others.

Listening source (lecture): after reading the text, participants listened to a lecture (3 min 26 sec) in which a professor introduced two studies illustrating examples of the audience effects.

Question: participants were asked to respond to a question regarding how the two examples provided in the lecture could illustrate the audience effects.

Reference:

Reading source:

Educational Testing Service. (2006). *The official guide to the new TOEFL iBT*. McGraw Hill. p. 231

Listening source:

Educational Testing Service. (2006). *The official guide to the new TOEFL iBT*. McGraw Hill. Disc 2: Track 27

## Appendix C

List of discourse features examined

|  |  |
| --- | --- |
| Signposting | Examples (possible variations) |
| **Exemplifiers**  (Introducing examples) | * For example * That’s an example of… |
| **Sequential marker**  (Indicating the sequence of information) | * the first (example/time/day/thing/group/experiment) * the second (example/situation) * (and/but/just) then * And after that |
| **Contrast**  (Contrasting two or more information) | * compared to * in comparison * in this/that (case/situation/experiment) * in case of * another example is * whereas * on the other hand |
| **Summarizer**  (Initiating a summarizing statement) | * In conclusion * And conclusion is that * So, this/that is how/what… |
| **Source attribution**  (Reference to reading and/or listening sources) | * in the (reading) passage * in the lecture * according to the lecture/research * according to the professor/lecturer/instructor |

|  |  |
| --- | --- |
| Connectives | Included items |
| **Disjunction** | *or* |
| **Addition** | *and, also, besides, further, furthermore, too, moreover, in addition, then, another, indeed, likewise* |
| **Causal connectives** | *although, arise, arises, arising, arose, because, cause, caused, causes, causing, condition, conditions, consequence, consequences, consequent, consequently, due to, enable, enabled, enables, enabling, even then, follow that, follow the, follow this, followed that, followed the, followed this, following that, follows the, follows this, hence, made, make, makes, making, nevertheless, nonetheless, only if, provided that, result, results, since, so, therefore, though, thus, unless, whenever* |
| **Opposition** | *but, however, nevertheless, otherwise, on the other hand, on the contrary, yet, still, maybe, perhaps, instead, except for, in spite of, despite, nonetheless, apart from, unlike, whereas* |

|  |  |
| --- | --- |
| Structural components | Description, information to be included |
| **Introduction** | General idea/topic of the speech |
| **Abstract idea/Concept explanation – 1st example** | Prompt A (psychology)  People tend to explain other’s behavior based on their personality (because they are unaware of others’ situation)  Prompt B (sociology)  People act faster/differently if they know they are observed by someone |
| **Abstract idea/Concept explanation – 2nd example** | Prompt A (psychology)  People tend to explain their own behavior based on situational factors (because we know situational forces around us)  Prompt B (sociology)  Certain common behaviour pattern (e.g., making mistakes) will also increase when people know they are being watched. |
| **Listening – 1st example** | Prompt A (psychology)  a man cut in line, and the professor felt he was rude  Prompt B (sociology)  people tie shoes faster when they know they are being observed compared to those who didn’t know they are being observed |
| **Listening – 2nd example** | Prompt A (psychology)  the professor cut in line, but he didn’t feel himself was selfish because he was in a hurry  Prompt B (sociology)  people type computers faster when they know they are being observed; but they also make more mistakes |
| **Summary** | Summary or repetition of the idea/topic of the speech. This is often introduced by “so…” “in conclusion…” |

## Appendix D

Rater background questionnaire

**Rater Background Questionnaire**

1. Gender: Female/Male/Other/prefer not to say

2. Age: \_\_\_\_\_\_\_ (years)

3. Are you currently a student? Yes/No

4. Current Degree/ Major/ Year of study (if applicable): \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(e.g. MA/ Applied linguistics/ 2nd year)

5. Last Degree you earned/ Major: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

6. Is your hearing normal as far as you know? Yes/No

7. If your hearing is not normal, please explain: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Language Use**

9. What is/are your native language(s) (from birth)? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

10. Which other languages do you know or speak at some level? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

11. Of these languages, which would you say that you are proficient in? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

12. For how long have you lived in Canada? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ years

13. Approximately what percent of the time do you speak **English** per day (as opposed to other languages)?

0% 100%

14. Approximately what percent of the time do you spend listening to **English** per day (as opposed to other languages)?

0% 100%

15. How familiar are you with following accented English?

**Chinese**

*Not at all familiar* *Very familiar*

**Spanish**

*Not at all familiar*  *Very familiar*

**French**

*Not at all familiar*  *Very familiar*

**Farsi**

*Not at all familiar*  *Very familiar*

**Hindi/Urdu**

*Not at all familiar*  *Very familiar*

Do you consider yourself as a second language speaker of English? (i.e., not a native speaker)

If yes, please answer following questions.

16. At what age did you start learning English? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

17. Where did you learn English? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

18. Using the scale below, please rate level of English proficiency in each of the following areas:

**Speaking**

Not at all proficient Very proficient

**Listening**

Not at all proficient Very proficient

**Reading**

Not at all proficient Very proficient

**Writing**

Not at all proficient Very proficient

**Experience in teaching and assessment**

19. Please tell us about your English teaching experience.

For example, you can describe: How long have you taught academic English?  How many course(s)? Which country(ies)? What was the school level? (e.g., CEGEP, university) What did you teach in the course(s)?

20. Please briefly explain what you have done in your class(es) to teach academic English speaking skills (e.g., pronunciation training, helping students making oral presentations, IELTS preparation).

21. Do you have any experience as a professional rater or examiner, such as for an institutional language placement test or standardized proficiency test (e.g., TOEFL, IELTS)?

If yes, please describe the context (e.g., IELTS speaking test examiner, 2 years)

## Appendix E

## Correlation matrix across discourse features

## Appendix F

A graph of a graph with dots

Description automatically generated with medium confidenceMarginal effects plot illustrating the effect of coherence on comprehensibility, with a linear function and 95% CI estimates (shaded in light blue) showing the best-fitting trendline

**Final model formula for full mixed-effects model for comprehensibility**:

comprehensibility ~ coherence + accentedness + prompt + speaker’s L1 + listener’s Mandarin-accent familiarity + listener’s Farsi-accent familiarity + listener’s Indian-accent familiarity + listener’s Romance-accent familiarity + (1|speaker) + (1|rater)

## Appendix G

Scatterplots plotted separately for speakers’ L1 backgrounds

Chart, scatter chart

Description automatically generated

**Appendix H**

**Final model formula for full mixed-effects model for coherence and discourse features**:

coherence ~ structure-ordering + signposting + reading source-speech similarity + listening source-speech similarity + disjunction + causal connectives + opposition + addition + overlap nouns + overlap verbs + pronoun-noun ratio + attended demonstrative + prompt + speaker’s L1 + (1|speaker) + (1|rater).

**Final model formula for full mixed-effects model for comprehensibility and discourse features**:

comprehensibility ~ structure-ordering + signposting + reading source-speech similarity + listening source-speech similarity + disjunction + causal connectives + opposition + addition + overlap nouns + overlap verbs + pronoun-noun ratio + attended demonstrative + prompt + speaker’s L1 + (1|speaker) + (1|rater).

**Appendix I**

Summary of full mixed-effects model for coherence and discourse features

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Parameter | *Estimate* | *SE* | *95% CI* | *t* | *p* |
| (Intercept) | 56.54 | 5.15 | [46.17, 66.91] | 10.99 | <.001 |
| Organization |  |  |  |  |  |
| Structure-ordering | 4.22 | 1.42 | [1.38, 7.06] | 2.97 | .004 |
| Signposting | 1.41 | 1.42 | [–1.42, 4.24] | 0.99 | .326 |
| Source-speech similarity |  |  |  |  |  |
| Reading | 4.55 | 1.91 | [0.74, 8.36] | 2.38 | .021 |
| Listening | 2.49 | 1.52 | [–0.55, 5.51] | 1.63 | .108 |
| Connectives |  |  |  |  |  |
| Disjunction | 0.58 | 1.66 | [–2.73, 3.89] | 0.35 | .727 |
| Causal connectives | 0.61 | 1.44 | [–2.26, 3.47] | 0.42 | .675 |
| Opposition | 0.97 | 1.55 | [–2.12, 4.06] | 0.63 | .534 |
| Addition | –1.17 | 1.40 | [–3.95, 1.61] | –0.84 | .407 |
| Lexical overlaps |  |  |  |  |  |
| Overlap nouns | 0.85 | 1.52 | [–2.18, 3.87] | 0.56 | .580 |
| Overlap verbs | 3.26 | 1.54 | [0.19, 6.33] | 2.12 | .039 |
| Givenness |  |  |  |  |  |
| Pronoun-noun ratio | 1.18 | 1.38 | [–1.57, 3.93] | 0.86 | .396 |
| Attended demonstrative | 1.41 | 1.55 | [–1.69, 4.50] | 0.90 | .370 |
| Control-covariates |  |  |  |  |  |
| Prompt (psychology vs. sociology) | –6.34 | 4.15 | [–14.61, 1.94] | –1.53 | .133 |
| Speakers’ L1 |  |  |  |  |  |
| Mandarin vs. Farsi | 7.27 | 4.00 | [–0.71, 15.24] | 1.82 | .264 |
| Mandarin vs. Indian | 7.94 | 3.83 | [0.31, 15.57] | 2.07 | .161 |
| Mandarin vs. Romance | 9.83 | 3.93 | [1.99, 17.65] | 2.50 | .059 |
| Random effects | *Variance* | *SD* | *Criterion* | *Estimate* | |
| Rater (intercept) | 161.26 | 12.70 | Log-likelihood | –3024.2 | |
| Speaker (intercept) | 55.99 | 7.48 | AIC | 6088.3 | |
|  |  |  | BIC | 6179.8 | |
|  |  |  | Marginal *R2* | 0.14 | |
|  |  |  | Conditional *R2* | 0.56 | |

*Note*. *p* values for speakers’ L1 were adjusted using Tukey correction

Summary of full mixed-effects model for comprehensibility and discourse features

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Parameter | *Estimate* | *SE* | *95% CI* | *t* | *p* |
| (Intercept) | 60.57 | 6.01 | [48.30, 72.84] | 10.09 | <.001 |
| Organization |  |  |  |  |  |
| Structure-ordering | 2.67 | 1.37 | [–0.06, 5.40] | 1.95 | .056 |
| Signposting | 0.13 | 1.37 | [–2.59, 2.85] | 0.09 | .925 |
| Source-speech similarity |  |  |  |  |  |
| Reading | 2.67 | 1.84 | [–0.99, 6.33] | 1.45 | .152 |
| Listening | 2.60 | 1.46 | [–0.32, 5.51] | 1.78 | .080 |
| Connectives |  |  |  |  |  |
| Disjunction | 0.74 | 1.60 | [–2.45, 3.92] | 0.46 | .646 |
| Causal connectives | –0.31 | 1.38 | [–3.06, 2.44] | –0.22 | .825 |
| Opposition | 0.63 | 1.49 | [–2.33, 3.61] | 0.43 | .672 |
| Addition | –2.72 | 1.34 | [–5.40, –0.05] | –2.03 | .047 |
| Lexical overlaps |  |  |  |  |  |
| Overlap nouns | 0.80 | 1.46 | [–2.11, 3.70] | 0.55 | .586 |
| Overlap verbs | 3.75 | 1.48 | [0.80, 6.70] | 2.53 | .014 |
| Givenness |  |  |  |  |  |
| Pronoun-noun ratio | 1.56 | 1.33 | [–1.08, 4.20] | 1.18 | .244 |
| Attended demonstrative | 1.00 | 1.49 | [–1.98, 3.98] | 0.67 | .506 |
| Control-covariates |  |  |  |  |  |
| Prompt (psychology vs. sociology) | –2.52 | 3.99 | [–10.48, 5.44] | –0.63 | .530 |
| Speakers’ L1 |  |  |  |  |  |
| Mandarin vs. Farsi | 9.39 | 3.85 | [1.72, 17.04] | 2.44 | .066 |
| Mandarin vs. Indian | 7.00 | 3.85 | [–0.34, 14.34] | 1.90 | .225 |
| Mandarin vs. Romance | 10.98 | 3.78 | [3.45, 18.50] | 2.91 | .020 |
| Random effects | *Variance* | *SD* | *Criterion* | *Estimate* | |
| Rater (intercept) | 288.21 | 16.98 | Log-likelihood | –3004.4 | |
| Speaker (intercept) | 51.49 | 7.18 | AIC | 6048.8 | |
|  |  |  | BIC | 6140.3 | |
|  |  |  | Marginal *R2* | 0.12 | |
|  |  |  | Conditional *R2* | 0.66 | |

*Note*. *p* values for speakers’ L1 were adjusted using Tukey correction